Final Technical Report

Second International Conference on Ordinal Data

ONR Grant N00014-94-1-0794

Principal Investigator: M. F. Janowitz

The conference was held October 14-17, 1993. There were a total of 59 participants from 10 different countries. The focus of the conference centered on the following topics:

1. What is the nature of ordinal data?
2. How does ordinal data arise and what are its uses?
3. What techniques have been used in the past to help interpret various types of ordinal data?
4. What directions should be taken by future research?

The conference was highly successful for a number of reasons. First of all, the number of attendees was small enough so that people got to know each other and had ample opportunity for both formal and informal discussions. This was further enhanced by the fact that several breaks were scheduled during the day, and that meals, residence, and the meeting rooms were all located in the same building. There were people from a variety of disciplines, and many different views of the subject emerged. The discussions were spirited but constructive, and many people have since written to say how successful they thought the conference was. Indeed, there are informal plans for holding a third such conference in France in 1996 under the sponsorship of INRIA, and a fourth conference in Russia in 1998 to be sponsored by the research group headed by Fuad Alekserov.

I am enclosing copies of the list of participants, the program for the meeting, and the abstracts. It should be noted that thanks to the funding provided by the Grant, we were able to attract some of the top names in the World who work in this area. Especially worthy of mention are:

Edwin Diday of France
J.-P. Doignon of Belgium
Pierre Hansen of Canada
Bernard Monjardet of France
Bruno Leclerc of France
Ivan Rival of Canada
Fred Roberts of Rutgers
Rudolf Wille of Germany

From the point of view of the research being supported by ONR Grant N00014-90J-1008, it is worthy of note that the conference gave a marvelous opportunity to showcase the work being done under this Gramt, and to attract the interest of both research mathematicians and journal editors in what is being done.

It should also be mentioned that there was a short course preceding the conference devoted to software and techniques developed in Darmstadt for processing ordinal data. This course attracted a number of people from Industry, as well as a sizable contingent from our Department of Sociology. A total of 19 people participated in the short course.

Participants in the Second International Conference on Ordinal Data Analysis

Department of Mathematics and Statistics
University of Massachusetts
Amherst, MA 01003, USA
October 15 – 17, 1993

Joseph L. Balloun
Department of Management and Marketing
Louisiana Tech University
P.O. Box 10318
Ruston, LA 71272, USA

Jean-Pierre Barthélemy
Laboratoire d'Intelligence Artificielle et Systems Cognitifs
Ecole Nationale Superieure des Telecommunications de Bretagne
BP 832-29285 BREST Cedex, FRANCE

M.K. Bennett
Department of Mathematics and Statistics
University of Massachusetts
Amherst, MA 01003, USA

Ulf Bockenholt
Department of Psychology
University of Illinois
603 East Daniel Street
Champaign, IL 61820, USA

Jon Breslaw
Department of Economics
Concordia University
1455 De Maisonneuve Blvd W.
Montreal, Que H3G 1M8, CANADA

Norman Cliff
Department of Psychology
University of Southern California
Los Angeles, CA 90089-1061, USA

Frank Critchley
School of Mathematics and Statistics
University of Birmingham
Birmingham, UNITED KINGDOM

Gary D. Crown
Department of Mathematics and Statistics
Wichita State Univeristy
Wichita, KS 67260-0033, USA

Deanna Dearholt
Department of Sociology
New Mexico State University
Las Cruces, NM 88003, USA

Don Dearholt
Department of Computer Science
Mississippi State University
Mississippi State, MS 39762, USA

Alfred P. De Fonso
Department of Electrical and Computer Engineering
University of Massachusetts
Amherst, MA 01003, USA

Edwin Diday
INRIA-Rocquencourt
Domaine de Voluceau
78153 Le Chesnay Cedex, FRANCE

Jean-Paul Doignon
Départment de Mathématique
Université Libre de Bruxelles
C.P. 216
Bld du Triomphe
1050 Bruxelles, BELGIUM

Sigrid Flath
Fachbereich Mathematik
Technische Hochschule Darmstadt
Schlossgartenstr. 7
64289 Darmstadt, GERMANY

David Foulis
Department of Mathematics and Statistics
University of Massachusetts
Amherst, MA 01003, USA

Herman Friedman
Apartment 4F
45 East End Avenue
New York, NY 10028, USA

Pierre Hansen
GERAD et Department des Methodes
Quantitative et Systemes d'Information
Ecole des Hautes Etudes Commerciales
5255 Avenue Decelles
Montreal, Quebec, CANADA H3T 1V6

Robert Jamison
Department of Mathematical Sciences
Clemson University
Clemson, SC 29634, USA

Melvin F. Janowitz
Department of Mathematics and Statistics
University of Massachusetts
Amherst, MA 01003, USA

Timo Koski
Department of Mathematics
Luleå University of Technology
S-971 87 Luleå, SWEDEN

Bruno Leclerc
E.H.E.S.S. - C.N.R.S. - PARIS-IV
Centre d'Analyse et de Mathématique Sociales
54, boulevard Raspail
75270 Paris Cedex 06, FRANCE

Fritz Lehmann
4282 Sandburg Way
Irvine, CA 92715, USA

Ingo Leonhardt
Fachbereich Mathematik
Technische Hochschule Darmstadt
Schlossgartenstr. 7
64289 Darmstadt, GERMANY

Andrey V. Malishevski
Institute of Control Sciences
Profsoyusnaya 65
Moscow 117806, RUSSIA

James McIntosh
Department of Economics
Concordia University
1455 De Maisonneuve Blvd W.
Montreal, Que H3G 1M8, CANADA

F. R. McMorris
Department of Mathematics
University of Louisville
Louisville, KY 40292, USA

Venu Menon
Department of Mathematics
University of Connecticut
Stamford, CT 06903, USA

Boris Mirkin
Department of Mathematics
Rutgers University
New Brunswick, NJ 08903-0270, USA

Bernard Monjardet
E.H.E.S.S. - C.N.R.S. - PARIS-IV
Centre d'Analyse et de Mathématique Sociales
54, boulevard Raspail
75270 Paris Cedex 06, FRANCE

Ken Paap
Department of Psychology
New Mexico State University
Las Cruces, NM 88003, USA

Piero Pagliani
Diresione Formasione ITASTEL
Via F. Depero 24
00155 Rome, ITALY

Aleksander Pekec
Department of Mathematics
Rutgers University
New Brunswick, NJ 08903, USA

Dale Peterson
Department of Mathematics
Rutgers University
New Brunswick, NJ 08903-5062, USA

Robert C. Powers
Department of Mathematics
University of Louisville
Louisville, KY 40292, USA

K. Brooks Reid
Mathematics Department
California State University
San Marcos, CA 92096-0001, USA

Ivan Rival
Department of Computer Science
University of Ottawa
Ottawa K1N 9B4, CANADA

Fred S. Roberts
Department of Mathematics
Rutgers University
New Brunswick, NJ 08903-2101, USA

Tammo Rock
Fachbereich Mathematik
Technische Hochschule Darmstadt
Schlossgartenstr. 7
64289 Darmstadt, GERMANY

Michael Roddy
Mathematics Department
Brandon University
Brandon, Manitoba, CANADA R7A 6A9

Til Schuermann
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974, USA

Berthold Schweizer
Department of Mathematics and Statistics
University of Massachusetts
Amherst, MA 01003, USA

Dr. Norbert Spangenberg
Sigmund-Freud-Institut
Mylinstr. 20
6000 Frankfurt, GERMANY

Ralph Stinebrickner
Department of Mathematics
Berea College
Berea, KY 40404, USA

John P. Sutcliffe
Department of Psychology
University of Sydney
Sydney, NSW 2006, AUSTRALIA

Mike Sutherland
Department of Mathematics and Statistics
University of Massachusetts
Amherst, MA 01003, USA

Scott Sykes
Department of Mathematics and Statistics
University of Massachusetts
Amherst, MA 01003, USA

Rudolf Wille
Fachbereich Mathematik
Technische Hochschule Darmstadt
Schlossgartenstr. 7
64289 Darmstadt, GERMANY

Uta Wille
Fachbereich Mathematik
Universität Giessen
Arndtstr. 2
35390 Giessen, GERMANY

Michaela Winterberg
Fachbereich Mathematik
Technische Hochschule Darmstadt
Schlossgartenstr. 7
64289 Darmstadt, GERMANY

Monika Wols
Fachbereich Mathematik
Technische Hochschule Darmstadt
Schlossgartenstr. 7
64289 Darmstadt, GERMANY

Shaoji Xu
Department of Mathematics
Rutgers University
New Brunswick, NJ 08903-5062, USA

Elias Zintsaras
Laboratory of Mathematical Biology
The Ridgeway, Mill Hill
National Institute for Medical Research
London, ENGLAND NW7 1AA

Additional Participants in the Short Course on Formal Concept Analysis:

Andy B. Anderson
Department of Sociology
University of Massachusetts
Amherst, MA 01003, USA

Tim Dion
Marble Associates
38 Edge Hill Road
Waltham, MA 02154, USA

Gene A. Fisher
Department of Sociology
University of Massachusetts
Amherst, MA 01003, USA

Zhonge Li
Department of Sociology
University of Massachusetts
Amherst, MA 01003, USA

Michael Oakes
Department of Sociology
University of Massachusetts
Amherst, MA 01003, USA

Patrick Rusk
Marble Associates
38 Edge Hill Road
Waltham, MA 02154, USA

A. Chase Turner
Marble Associates
38 Edge Hill Road
Waltham, MA 02154, USA

Second International Conference on Ordinal Data Analysis

Department of Mathematics and Statistics
University of Massachusetts
Amherst, MA 01003, USA

Organizing Commitee:

Melvin F. Janowitz (Amherst, Massachusetts)
Ivan Rival (Toronto, Ontario)
Fred Roberts (New Brunswick, New Jersey)
Selma Strahringer (Darmstadt, Germany)
Rudolf Wille (Darmstadt, Germany)

Local Contact:

Professor Melvin F. Janowitz
Department of Mathematics and Statistics
University of Massachusetts
Amherst, MA 01003, USA

Phone: (413)-545-2871
E-mail: ordinal@math.umass.edu

General Information:

The conference will take place on the 9th floor of the Campus Center. Plenary lectures and Session I will be in Room 917, Session II in Room 911. Coffee and refreshments are served during the breaks in Room 917.

Registration:

Thursday: 8.30 – 9.30 a.m. in front of Room 162 of the Campus Center
Friday: 8.00 – 9.00 a.m. in the 9th floor lobby of the Campus Center

Preliminary Events:

Thursday, October 14

Short Course on Formal Concept Analysis held by Rudolf Wille (Darmstadt, Germany)	
9.00 – 1.00 p.m.	in Room 162 of the Campus Center
2.30 – 5.00 p.m.	in the PC-Lab, Room 01 (Basement) of the School of Management (SOM)
7.00 – 9.00 p.m.	WELCOMING RECEPTION IN ROOM 162 OF THE CAMPUS CENTER

Saturday, October 16

Time	Section I	Section II
	Chair: Mary K. Bennett	
9.00	**David J. Foulis** (Amherst, Massachusetts) Test Spaces and Data Acquisition	
9.45	**Uta Wille** (Giessen, Germany) Geometric Representation of Ordinal Contexts	
10.30	COFFEE BREAK	
	Chair: Fred S. Roberts	
11.30	**Jean-Paul Doignon** (Brussels, Belgium) Unimodality and Straightness of Multipreference Data	
12.15	**Don Dearholt** (Starkville, Mississippi) An Introduction to Pathfinder Networks, Proximity Graphs and Applications	
1.00	LUNCH FOR THE PARTICIPANTS IN ROOM 1001	
	Section I	Section II
	Chair: Don Dearholt	*Chair: Bernhard Monjardet*
2.30	**Ken Papp** (Las Cruces, New Mexico) Predicting Human Memory and Judgement Time from Ordinal Measures of Psychological Proximity	**Andrey V. Malishevski** (Moscow, Russia) Generalizations of Transitivity for Relations Between Sets
3.00	**John P. Sutcliffe** (Sydney, Australia) Weight of Evidence for Strict Simple Ordering in Pair Comparison Data Subject to Error	**Norman Cliff** (Los Angeles, California) Ordinal Alternatives to Mean Comparisons
3.45	COFFEE BREAK	
	Section I	Section II
	Chair: Berthold Schweizer	*Chair: Jean-Paul Doignon*
4.15	**Norbert Spangenberg** (Frankfurt, Germany) Implicit Knowledge of Psychoanalysis Concerning the Suitability of Patients for Psychotherapy	**James McIntosh** (Montreal, Quebec) Multivariate Analysis of Ordered Categorical Data
4.45	**Ingo Leonhardt, Michaela Winterberg** (Darmstadt, Germany) CONCLASS – CONceptual CLustering ASSistant	**Robert Jamison** (Clemson, South Carolina) Embedding Graphs in Disjunctive Products of Lines
5.15	**Tammo Rock** (Darmstadt, Germany) TOSCANA'1 – A Management System for Conceptual Data	**Piero Pagliani** (Rome, Italy) A Logic-Algebraic Approach to Formal Contexts with Partial Information

Second International Conference on Ordinal Data Analysis

Department of Mathematics and Statistics
University of Massachusetts
Amherst, MA 01003, USA

October 15 – 17, 1993

Abstracts of the Lectures

Second International Conference on Ordinal Data Analysis. University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

Solving the Complete Model of the Law of Comparative Judgment

Joseph L. Balloun

Dept. of Management and Marketing
Louisiana Tech University
P.O. Box 10318
Reston, LA 71272, USA

This paper presents a numerical procedure for estimating the latent parameters of the Thurstonian law of comparative judgment for data obtained by a common experimental procedure. For commonly encountered sample sizes, jackknife analyses allow estimates of the standard errors and confidence limits for the estimated parameters, and the overall goodness-of-fit function may be tested.

Past approaches to solving the Thurstonian law of comparative judgments have encountered two sets of challenges. First, the complete model involves more unknowns to be estimated than the number of observed equations available from experimental data. Second, the experimental judgment procedures commonly used yield restricted information about the objects ("stimuli") to be measured, under conditions not readily generalizable to the situation where the stimuli will be used for measurement.

There are at least three strategies for reducing the two major barriers to solving the complete Thurstonian law of comparative judgments. First, past efforts have restricted the effective number of parameters to be estimated by making restrictive assumptions about their values. Second, it is possible to wrest more observed equations from experimental data. Finally, a different experimental design may yield more information (direct estimates or more equations) about the latent parameters.

In the social sciences, a common experimental procedure, and the one assumed in this paper, is to have people respond to several different stimuli which are intended to measure one construct or concept. Examples include attitude surveys or academic tests. Such a design allows direct estimation of the correlations among people or among stimuli. Moreover, paired comparisons between people and between stimuli can be computed from the observed data matrix. The marginal means of people and stimuli provide additional observation equations. The various first derivatives of the observed equations relative to the remaining unknown parameters can be found, and a conjugate gradient method can be used to estimate the latent locations and "discriminal dispersions" of the stimuli and the people in a common space.

The proposed method is demonstrated with an artificial data set, and with data from an academic intelligence test. Some advantages and challenges to the experimental method and to the numerical analysis method are discussed.

Second International Conference on Ordinal Data Analysis, University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

Representation of Binary Data by Median Semilattices

Jean-Pierre Barthelemy

Laboratoire d'Intelligence Artificielle et Systems Cognitifs
Ecole Nationale Superieure des Telecommunications de Bretagne
BP 832-29285 BREST Cedex, FRANCE

A Guttman scale allows for ordering objects (or items or stimuli) described by "presence-absence" attributes (only one value: "presence" is taken into account). We study some generalizations of Guttman scales in the case of binary data (the two modalities have to be considered). We emphasize the use of median semilattices to represent such data.

Second International Conference on Ordinal Data Analysis, University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

Multivariate Analysis of Ordered Categorical Data

Jon A. Breslaw

Department of Economics
Concordia University
1455 De Maisonneuve Blvd W.
Montreal, Que H3G 1M8, CANADA

The problem considered in this paper is how to obtain both consistent and efficient estimates of the parameter vector β in the stochastic regression model,

$$y^* = x^*\beta + \mu$$

when only categorical responses, $\{y, x\}$, are available for all of the variables. When x^* is observable the appropriate technique is the standard ordered probit model. When only x is available, replacing the exogenous variables by dummies, as was done in McIntosh *et al* (1989), is not satisfactory. Furthermore, the "latent variable approach" has proved to be intractable when the matrix of polychoric correlation coefficients is not positive definite. Thus there is, at present, no technique which adequately deals with the multivariate nature of the problems.

Our approach generalizes a two-stage procedure proposed by Ross and Zimmermann (1987) using the conditional moments of x^*. The first stage involves the estimation of the matrix of delineation points and the variance-covariance matrix for the distribution of x^*. A one-factor analytic multivariate normal distribution is employed, and this permits the easy and efficient estimation of the distributional parameters even when the number of explanatory variables is greater than three. Using conditional means generated from the first stage, a semi-parametric method is used in the second stage to estimate the β parameters. The method is applied to a sample of Canadian firms which replied to a Conference Board of Canada survey questionnaire on investment intentions.

Ordinal Alternatives to Mean Comparisons

Norman Cliff

Department of Psychology
University of Southern California
Los Angeles, CA 90089-1061, USA

An obvious alternative to comparing the location of two distributions by comparing their means is to use a procedure based on ranks. A widely known procedure that does not require that the distributions be the same except for location is the delta measure, using the Zaremba-Fligner-Policello estimate of the variance of the estimate of delta, but it is restricted to independent groups data. However, the variance of the delta estimate in the case of paired observations can be defined and used as the basis for ordinal inferences with a repeated measure, alternative to the paired-observations t-test. It has certain conceptual and statistical advantages over statistics such as the sign test and the Wilcoxon signed-ranks test. Some examples are given,and a graphical aid to interpretation of these statistics, the "dominance diagram", is suggested for both independent groups and paired observations cases. Ways in which these location-comparison procedures can be generalized to more complex designs are suggested.

In a separate study, the size, power, and coverage of the Zaremba-Fligner-Policello method for making inferences about the independent groups delta measure is being investigated. Some problems with coverage in special cases have arisen due to the correlation between the estimate d and its estimated variance: when delta happens to be overestimated, its variance is underestimated, and a symmetric confidence interval tends to be too short on the side nearer zero. A modification that improved coverage was found. It is suggested these problems are similar to those that occur with other bounded measures and that they may occur rather widely with ordinal statistics when confidence intervals are used.

Second International Conference on Ordinal Data Analysis. University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

On Level Maps

Frank Critchley

School of Mathematics and Statisti:s
University of Birmingham
Birmingham, UNITED KINGDOM

A series of classical papers in the 1960's, notably Benzécri (1965), Hartigan (1967), Jardine, Jardine and Sibson (1967) and Johnson (1967), established that the ultrametrics of a finite set S can be put in one-to-one correspondence with either the dendograms on S or the indexed hierarchical classifications on S. Several extensions of these bijections have been given in the literature. In particular, Jardine and Sibson (1971) generalise from the ultrametrics to any (symmetric) dissimilarity; Janowitz (1978) and Barthélémy, Leclerc and Monjardet (1984) further allow the dissimilarity to take values not only in the nonnegative reals but in any join semi-lattice with a smallest element; and, most recently, Critchley and Van Cutsem (1989) deal with nonnegative-valued dissimilarities which are not necessarily symmetric (called predissimilarities) and which are defined on an arbitrary set S.

We review this field and show how one can in fact produce a single general theorem of bijection between the predissimilarities on an arbitrary set S taking values in an arbitrary ordered set L and certain increasing families of subsets of $S \times S$ indexed by L. Indeed this theorem is in turn the special case $E = S \times S$ of a natural dual order-isomorphism, for an arbitrary set E, between all functions $f : E \longrightarrow L$ and all so-called "level" maps $m : L \longrightarrow \mathcal{P}(E)$. Note that there are no constraints on the ordered set L. In particular, L need to be totally ordered.

The theory and applications of level maps are briefly indicated. Several alternative characterisations of level maps are presented. Some of these involve conditions on ordered sets that appear to be original. Their relationship with residual maps is explored. Applications include certain natural generalisations of standard methods of dissimilarity analysis. In particular, to asymmetric and/or multiattribute data.

Second International Conference on Ordinal Data Analysis, University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

Morphisms of Stability Families

Gary D. Crown

Department of Mathematics and Statistics
Wichita State Univeristy
Wichita, KS 67260-0033, USA

Many authors have employed the idea of a *stability family* to study various kinds of consensus methods. Let X be a set on which a consensus method is to be applied. A stability family on X is a pair (S, γ) where $\gamma : X \longrightarrow \mathcal{P}(S)$ is an injective function from X to the set of subsets of S. One can think of the set $\gamma(x)$ as a set of building blocks for the element $x \in X$. Let $x, y \in X$. Define $x \leq y$ if and only if $\gamma(x) \leq \gamma(y)$. If V is a finite set of voters, then a consensus function on X is a function $F : X^V \Longrightarrow X$. We investigate how the possible consensus methods vary with different stability families which induce the same order on X. To facilitate this investigation, various types of morphisms of stability families are introduced and properties of consensus functions which are preserved by these morphisms are studied.

Second International Conference on Ordinal Data Analysis. University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

An Introduction to Pathfinder Networks, Proximity Graphs and Applications

Don Dearholt

Dept. of Computer Science
Mississippi State University
Mississippi State, MS, USA

The Pathfinder network model for human semantic memory is presented, and the motivations, approaches, and research objectives of the paradigm are discussed. Parameters controlling the density of edges in Pathfinder networks are described, and the control of violations of the metric axioms are discussed. Theoretical results include the construction of two orthogonal families of networks associated with a given dissimilarity matrix, the capability of generating single-link hierarchical clustering from any Pathfinder network (the converse is not true), the preservation of edge structure under multiplicative transformations of the dissimilarity measures, and the preservation of edge structure under monotonic transformations (preserving order) when the r-metric is infinity.

Proximity graphs, which require a coordinate system, are briefly introduced, and motivations and applications are mentioned. Relationships between Pathfinder networks and some proximity graphs are given, and a new family of proximity graphs derived from Pathfinder networks is discussed.

An overview of some applications of Pathfinder networks and of their extensions is given. These include a hypertext browser for an online UNIX help system, a robotic vision database system, and modeling dynamic systems using the co-occurrence of entities. The latter results in three levels of representation: unconstrained by either partitional or hierarchical constraints, constrained by the generalized triangle inequality (the co-occurrence Pathfinder network), and the hierarchical constraints resulting in single-link clustering. Some situations in which the middle level of representation (the co-occurrence Pathfinder network) is preferable are discussed, and aspects of learning and consensus are presented in this context. Cumulative consensus is defined and illustrated using co-occurrence Pathfinder networks and the scenarios which can be used to generate them.

Second International Conference on Ordinal Data Analysis, University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

An Introduction to Symbolic Data Analysis

Edwin Diday

INRIA-Rocquencourt
Domaine de Voluceau
78153 Le Chesnay Cedex, FRANCE

The main aim of the symbolic approach in Data Analysis is to extend problems, methods and algorithms used on standard data to more complex data where the units are called "symbolic objects", in order to distinguish them from objects (described by numerical or categorical variables) treated by standard Data Analysis methods. "Symbolic objects" extend classical objects of data analysis at least in two ways: first, in the case of individuals, by giving the possibility of introducing structured information in their definition; second, in the case of sets or classes, by being "intensionally" defined. In both cases, in order to represent uncertain knowledge, it may be useful to use probabilities, possibilities (in case of vagueness and imprecision for instance), belief (for probabilities only known on parts and to express ignorance); that is why we introduce several kinds of symbolic objects: boolean, possibilist, probalilist, and belief. We briefly present some of their qualities and properties; three theorems, show how Probability, Possibility and Evidence theories may be extended to these objects. Some mixture decomposition problems on these objects are settled. We show that, in some cases, fractals are well adapted to representing duality between symbolic objects. Sets of symbolic objects are represented by categories of different kinds (hierarchies, pyramids, and lattices). Four kinds of data analysis problems including the symbolic extension are illustrated by several algorithms which induce knowledge from classical data or from a set of symbolic objects. Finally, important steps of a symbolic data analysis are described and illustrated by an example concerning road accidents.

Second International Conference on Ordinal Data Analysis, University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

Testing Unimodality and Straightness of Multipreference Data

Jean-Paul Doignon

Department of Mathematics
Universite Libre de Bruxelles
c.p. 216
Bd. du Triomphe
1050 Bruxelles, BELGIUM

Suppose that the objects forming a finite set X are ranked according to personal preference by the subject from a finite population A. Without loss of generality, it is assumed that the (simple) orders $\leq_a$ thus produced by the subjects a from A are all distinct. Then M= $(X, \{\leq_a | \; a \in A\})$ constitutes a *multipreference system.*

The unfolding model of Coombs [3] (see also Falmagne and Doignon [5]) offer an explanation of such data on the basis of relative proximity in some embedding 'psychological' space. Formally, any multipreference system M admits an Euclidean *r-representation* in the following sense: there exist two mappings f on A and g on X, both to some r-dimensional Euclidean space E^r, and such that for $a \in A$ and $i, j \in X$:

$$i \leq_a j \quad \Longleftrightarrow \quad |f(a) - g(i)| \leq |f(a) - g(j)|.$$

Several computer programs are available for tentative construction of Euclidean r-representations with lower values of r (see e.g. Borg and Lingoes [2]). However, given any r in $\{1, 2, \ldots\}$, no necessary and sufficient conditions seems to be known for the existence of an Euclidean r-representation.

A recent paper by Doignon and Falmagne [4] describes a polynomial-time algorithm that tests whether a multipreference system admits a 1-representation. The basic idea is to first check that two necessary conditions, called 'unimodality' and 'straightness', are fullfilled by the system. When they are, the problem is then reduced to an existence problem for a finite set of linear inequalities.

The talk will focus on these two conditions and the related sub-algorithms. Unimodality plays also a role in voting theory. The algorithm we present to test unimodality is better than a previous, indirect one by Bartholdi and Trick [1], at least for worst-case execution time, and also for insight in the concept itself. The notion of straightness will be introduced in a nice, geometric setting, by looking at the so-called permutohedron. The resulting algorithm is then compared with work done by Romero [6].

References:

[1] J. Bartholdi & M. A. Trick: *Stable matchings with preferences derived from a psychological model*, Operations Research Letters 5(4), 1986, 165-189.

[2] I. Borg & J. Lingoes: *Multidimensional Similarity Structure Analysis*. Springer Verlag, New York, 1987.

[3] C. H. Coombs: *A Theory of Data*. Wiley, New York, 1964.

[4] J.-P. Doignon & J.-Cl. Falmagne: *A Polynomial Time Algorithm for Unidimensional Unfolding Representations*. Journal of Algorithms, to appear.

[5] J.-Cl. Falmagne & J.-P. Doignon: *Bisector Spaces: Geometry for Triadic Data*. In: Mathematical Psychology: Current Developments (J.-P. Doignon & J.-Cl. Falmagne, eds.), pp. 89-105, Springer-Verlag, New York, 1991.

[6] D. Romeo: *Variations sur l'effet Condorcet*. Unpublished doctoral dissertation, Universit[illegible]n-tifique et Médicale de Grenoble, 1978.

Second International Conference on Ordinal Data Analysis, University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

Test-Spaces and Data Acquisition

David Foulis

Department of Mathematics and Statistics
University of Massachusetts
Amherst, MA 01003, USA

Test spaces generalize the standard notion of a sample space, but provide for situations in which there are several tests or measurements, some of which may be incompatible. Two tests are incompatible if they cannot be performed simultaneously without interfering with each other. The formal defintion of a test space is deliberately made as general as possible to accomodate a wide range of applications. Data acquired from executing the tests are provided with structure inherited from the patterns formed by the test-patterns which become significant for problems of statistical inference.

Second International Conference on Ordinal Data Analysis, University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

Cluster Analysis: A Point of View for Applications

Herman P. Friedman

Apartment 4F
45 East End Avenue
New York City, NY 10028, USA

Cluster Analysis is a term used to describe loosely connected sets of concepts criteria and algortithms for discovering and defining subgroups in observational data.

There are different ways of looking at methods of Cluster Analysis. One is to understand the individual methods. Another is to understand the process of using the methods in problem solving. The purpose of this talk is to focus on the use of cluster analysis in concert with related methods of data analysis in problem solving.

At present there is limited theory and formalism for directing the process of using clustering and related methods of data analysis in practice. There is, however, a reasonable consensus that any rational approach to applications should consider: purpose, choice of objects, type of data structure, nature of the problem, selection, transformation and *scaling of variables*, *choice of algorithm*, *criteria* of evaluation, and appropriate communication of applicability of the proposed classifications and models.

Suggestions for using the context of problems to guide the user in making choices within this frame work will be provided. Recommendations for software and computing environments will be given.

Second International Conference on Ordinal Data Analysis, University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

Sequential Single Cluster Clustering with Ordinal Data

Pierre Hansen

GERAD et Department des Methodes
Quantitative et Systemes d'Information
Ecole des Hautes Etudes Commerciales
5255 Avenue Decelles
Montreal, Quebec, CANADA H3T 1V6

Sequential Single Cluster Clustering proceeds, given a set of entities, by finding a subset (or cluster) of them which is homogeneous and/or well seperated, removing the entities of that cluster and iterating until no further cluster can be found. In the case of ordinal data (i.e., when only a ranking of the dissimilarities between pairs of entities is given) a criterion of homogeneity is the *split* of the cluster, or minimum dissimilarity between an entity inside this cluster and one outside of it; criteria for homogeneity are (i) the *radius* of the cluster, or maximum dissimilarity between an entity of the cluster and its center (also an entity, chosen to minimize the value of the radius) and (ii) the *diameter* of the cluster, or largest dissimilarity between a pair of entities of this cluster. Algorithms are proposed (or, in one case, reviewed) to maximize split, minimize radius or diameter for all possible cardinalities of the cluster, as well as for the bicriterion problems with split and radius and with split and diameter criteria. Detection of clusters from lists of candidate ones is discussed. Computational experience with test problems from the literature is reported.

Second International Conference on Ordinal Data Analysis, University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

Embedding Graphs in Disjunctive Products of Lines

Robert Jamison

Dept. of Mathematics
Clemson University
Clemson, SC, USA

The disjunctive product of lines, introduced by Strahringer and Wille, provides a notion of "betweenness" for ordinal data in several variables. This notion of betweenness leads to a new and rather strange notion of convexity on euclidean space. The embedding of graphs into this structure represents an attempt to capture the 1-skeleta of polytopes in this convexity. I will discuss several Steinitz type theorems about the kinds of graphs which can occur as polyhedral 1-skeleta.

Ordinal Approaches to Cluster Analysis

Melvin F. Janowitz

Departement of Mathematics and Statistics
University of Massachusetts
Amherst, MA 01003, USA

The goal of cluster analysis is to suggest the possible internal structure of a data set by considering various attributes of the individual objects under consideration. The usual paradigm converts the attribute data into a measure of dissimilarity (DC), and then converts this intermediate DC into either a partition or a nested sequence of partitions of the objects. Of course if the original attributes have only ordinal significance, it is hard to image that any resulting DC will have more than ordinal significance. But even if the attributes are of scalar or interval type, the intermediate DC often still has only ordinal significance. Examples of this include the use of product-moment correlation or the cosine of the angle between the attribute vectors as measures of dissimilarity.

Ordinal clustering algorithms may be naturally classified according to how they interact with order preserving mappings of the nonnegative reals. This leads to a formal context whose structure is being investigated in joint work with Rudolf Wille. It can also be argued that the appropriate cluster algorithms for use with ordinal data (the so-called monotone-equivariant cluster methods) are graph-theoretic in nature. In joint work with Ralph Stinebrickner, it has been shown that a useful and interesting classification of such clustering algorithms may be achieved from pure graph-theoretic considerations, thus avoiding any interaction with any mappings on the reals. A brief introduction will be provided to both of these approaches.

Second International Conference on Ordinal Data Analysis. University of Massachusetts. Amherst. MA 01003. October 15-17, 1993

Methods in Probabistic Numerical Identification of Bacteria

Timo Koski

Dept. of Mathematics
Luleå Univ. of Technology
S-951 87 Luleå, SWEDEN

Probabilistic numerical identification of bacterial strains can be seen as joint estimation of and classification by a *finite mixture of multivariate Bernoulli distributions*, [1, 3]. In addition the number of classes occurring in the mixture has to be determined from data. The ***identification*** of a strain $x = (x_i)_{i=1}^{d}$ with binary features x_i, means assigning x to the class c_j that maximizes the *a posteriori* probability $P(c_j \mid x)$ estimated prior to or simultaneously with identification from data. Due to a subsequently increased chance of error the identification is combined with a *rejection option* known as the *Willcox* probability, see [3]. We analyse, c.f. [2], the performance of this kind of techniques by means of error-reject curves and discuss the problem of identifiability.

References

[1] M. Gyllenberg, T. Koski, M. Verlaan, E. Reilink: Nonuniqueness in Probabilistic Numerical Identification of Bacteria. *J. Appl. Prob.*, to appear, 1994.

[2] E. Reilink: *Evaluation of decision-directed learning regarding finite mixtures of multivariate Bernoulli distributions by means of error-reject curves*. Tech. Report, Luleå Univ. Techn., 1993.

[3] W.R. Willcox, S.P. Lapage, B. Holmes: A Review of Numerical Methods in Bacterial Identification. *Antonie van Leeuwenhook*, 46, pp. 233-299, 1980.

Second International Conference on Ordinal Data Analysis, University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

Generalizations of Transitivity for Relations between Sets

Andrey V. Malishevski

Institute of Control Sciences
Profsoyuznaya 65
Moscow 117806, RUSSIA

We deal with binary relations given on the power set of some basic set (finite for simplicity); call them **Hyperrelations**. Apart from the usual notion of the transitivity applied to such kind of relations, we consider stronger versions of similiar properties, calling them **Hypertransitivity**. These versions take into account the set-theorectical structure of objects connected by hyperrelations. Such generalizations are useful, e.g., in models of logical inference or dependencies in databases where one can infer new statements taken seperately. Equivalent representations of hypertransitive hyperrelations are suggested which use auxiliary families of representing sets or, alternatively, families of orderings on the basic set. These representations actually are a kind of generalization of the well-known Szpilrajn-Dushnik-Miller representation of usual transitive relations (preorders or partial orders). The obtained results are applied to elucidate mechanisms of generating some important types of set-to-set mappings. Two opposite cases are investigated: 1) expansions, with the axioms of abstract closure operators, and 2) contractions, with the so-called Path Independence Axiom originated in decision theory as applied to choice functions. Universal representations for both of these cases are obtained. The relationships with the Euler-Venn diagrams and with abstract convex geometries (antimatroids) are established as well.

Second International Conference on Ordinal Data Analysis, University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

Mathematical Classification and Clustering

Boris Mirkin

Central Economics - Mathematics Institute
Moscow 117418, Krasikova 32, RUSSIA

Mathematical classification theory should include the following subjects: classification structures; models connecting classification structures with data; algorithms to fit the models; interrelations between those models and algorithms; interpretation and knowledge generation; interrelations with other mathematical disciplines, which will be illustrated with an example based on an approximation clustering model. The paper will include a brief discussion of each of the above topics.

Second International Conference on Ordinal Data Analysis, University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

Latticial Theory of Consensus: An overview

Bernard Monjardet

Centre d'Analyse et de Mathematiques Sociales
54 Blvd Raspail
75270 Paris CEDEX 06, FRANCE

The consensus problem met in various fields (social choice, paired comparisons methods, cluster analysis, data analysis ...) is to "aggregate" an n-tuple of objects into a — or several — consensus object(s). Since the Arrow axiomatic approach and theorem (1951-1957) in the field of social choice, many "arrowian" results have been obtained in this field or other fields. Three years ago [5] I presented the foundations of a consensus latticial theory able to account of many of the results obtained when one want a unique consensus object (for a similar approach in a case where one can get several consensus objects (see [1]).

In this talk I first give the general framework, results and recent developments (especially in the case of valued objects, [3], [4]) of this theory; the properties of a *dependence* relation δ defined on the set of all join-irreducible of any (semi)lattice plays a key role in the obtaining of several of these results; then I show that this relation is linked with the *weak perspectivity* relations (or *arrows relations*, see [6]) defined between the join and the meet irreducible elements of the lattice and I give a report of the known properties of the dependence relation in some classes of lattices: finally I consider the generalization of this theory to an ordinal context and the connections with a recent work in this direction [2].

References:

[1] J. P. Barthelemy & M. F. Janowitz: *A formal theory of consensus.* SIAM J. Discr. Math. 4, 1991, 305-322.

[2] G. D. Crown, M. F. Janowitz & R. C. Powers: *Neutral Consensus Functions.* Mathematical Social Sciences, **25**, 3, 1990, 231-250.

[3] Leclerc, B.: *Aggregation of fuzzy preferences: a theoretic Arrow-like approach.* Fuzzy Sets and Systems **43**, 3, 1991, 291-309.

[4] Leclerc, B. & Monjardet, B.: *Latticial theory of consensus.* Proceedings of the First meeting of the Society for Social Choice and Welfare, to appear 1993.

[5] Monjardet, B.: *Arrowian characterizations of latticial federation consensus functions.* Mathematical Social Sciences, **20**, 1, 1990, 51-71.

[6] Wille, R.: *Subdirect decomposition of concept lattices.* Algebra Universalis **17**, 1983, 275-287.

Predicting Human Memory and Judgment Time from Ordinal Measures of Psychological Proximity

Kenneth R. Paap

Deparment of Psychology & Computing Research Laboratory
New Mexico State University
Las Cruces, NM 88003, USA

The organization of knowledge and the impact of this organization on memory and judgment has been of interest to cognitive psychologists for a long time. A classical way of representing knowledge has been through the use of networks where concepts are represented as nodes in a graph structure, and the relation between concepts as links between nodes. The Cognitive Systems Group at New Mexico State University (Schvaneveldt, 1990) has developed a technique called "Pathfinder" that empirically derives network structures from proximity data such as relatedness ratings. Pathfinder connects two items if all other network distances between those two items are greater than or equal to the distance estimate for that pair. Two parameters, r and q, determine how network distance is calculated and affect the density of the network. The parameter r defines path length in terms of the Minkowski metric and when $r = \infty$, path length equals the maximum weight of the links that make up the path. Thus, for $r = \infty$ only ordinal assumptions need to be made about the data.

The goal of this talk will be to review some of the recent evidence that supports the psychological validity of Pathfinder networks as a representation of human semantic memory. For example, performance on a cued recall task is influenced by whether the cue and target are linked directly or through a path involving a second node (Bajo & Canas, 1992). Performance is also affected by the number of direct links emanating from a cue node. These effects occur even when the cue-target pairs are equated for associative strength and relatedness ratings. In another recent study Cooke (1992) has shown that Pathfinder proximities predict response time in category and size judgments tasks better than the ratings from which the networks were derived and better than proximities obtained from multidimensional scaling. These successes may be attributed to Pathfinder's emphasis on relatedness estimates from the related end of the scale that tend to be more sensitive to differences in conceptual proximity than those at the unrelated end (Roske-Hofstrand & Paap, 1990). Pathfinder discards these least sensitive estimates and reconstructs them using the most sensitive estimates.

A Logic-Algebraic Approach to Formal Contexts with Partial Information

Piero Pagliani

Direzione Formazione ITASTEL
Via F. Depero 24
00155 Rome, ITALY

Burmeister has studied a logical model allowing the *calculus of dependency relations* among attributes in partial binary contexts, that is in contexts $\mathbb{K} = \langle G, M, I\rangle$, in which the information function I is just partially defined over the direct product $G \times M$.

Since these contexts are only partially informed, not only definite questions like "does A depends on B" ($A, B \in \wp(M)$) make sense, but also "not deterministic" questions like "the dependency of A and B is True or Not Definite, but not False" may be of relevance.

In my contribution, I provide a logic-algebraic model for the calculus of attribute dependencies in partial binary contexts. I show that any suitable interpretation of the notion of dependency among sets of attributes has to preserve some minimal logical requirements: namely the *compatibility* with Classical Logic and a *stability* condition. These requirements lead, from the algebraic point of view, to Kleene algebras and from a logical point of view to the Strong Semantics of Kleene. But in order to be able to obtain non-deterministic information about dependencies, we have to extend Kleene algebras to semi-simple Nelson algebras in order to be able to apply the "extended" implication operations $\prec$ and $\sqsupset$ that I introduce in order to supply the required non determistic information about dependencies.

Nevertheless, via a new intuitive model, that I call the "conveyed information model", it is shown that in a sense the Kleene logic is the natural logic connected to partial contexts.

It is interesting that an intuitive analysis of partial contexts provides in this way a "concrete" interpretation of the dual spaces of Kleene algebras.

More precisely, given a partial context $\mathbb{K} = \langle G, M, I\rangle$, I obtain a structure $\langle \mathrm{G}, \lesssim, f\rangle$, the *spectral space*, *Spect*($\mathbb{K}$), of the partial context $\mathbb{K}^* = \langle G^*, M^*, I^*\rangle$, where G is a set of elements which represent the general information behaviour of the original elements from G, the relation $\lesssim$ represents possible transmission of information from more defined elements of G to less informed elements and f is a track linking what is known about an element of G to what has not (yet) been defined.

The structure $\langle \mathrm{G}, \lesssim, f\rangle$ reveals to be the dual space of the Kleene algebra produced by the application of an operator π to the elements of M^* (that represent the general information behaviour of the original elements from M):

$$\pi(\{m\}) = \langle g \in G^* \text{ definitely enjoying } m \in M^*\}, \{g \in G^* \text{ definitely not enjoying } m \in M^*\}\rangle$$

In support of these results there is the fact that using different technical and intuitive approaches I have obtained essentially the same conclusions as those of Burmeister where the topic was originally studied.

Second International Conference on Ordinal Data Analysis, University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

Consensus Methods for n-Weak Hierarchies

Robert C. Powers

Department of Mathematics
University of Louisville
Louisville, KY 40292, USA

Let $n \geq 3$ be a positive integer. Bandelt et al. (1991) defined an **n-weak hierarchy H** on a finite set S to be a collection of subsets of S such that for any $A_1, \ldots, A_n$ in H there exists an index j such that

$$\bigcap_{i=1,\ldots,n} A_i = \bigcap_{\substack{i=1,\ldots,n \\ i \neq j}} A_i.$$

It was observed by Bandelt (1992) that a 3-weak hierarchy is a generalization of a prepyramid (see Diday (1986)).

A k-tuple of hierarchies $(H_1, \ldots, H_k)$ is called **mixed** if there is a decomposition $\{A_j : j = 3, \ldots, n\}$ of the index set $\{1, \ldots, k\}$ such that $i \in A_j$ implies H_i is a t-weak hierarchy for $t \geq j$. The problem of giving a well-defined method of consensus using k-tuples of mixed hierarchies as input and n-weak hierarchies as output is investigated.

References

H.-J. Bandelt (1992): Four-point characterization of dissimilarity functions obtained from indexed closed weak hierarchies, preprint.

H.-J. Bandelt, M. F. Janowitz, G. C. Meletiou (1991): n-median semilattice, *Order* 8, 185-195.

E. Diday (1986): Orders and overlapping clusters by pyramids, *Multidimensional data analysis Proc.*, eds. J. De Leeuw et al. DSWO Press, Leiden.

Second International Conference on Ordinal Data Analysis, University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

Drawing, Order and Surfaces

Ivan Rival

Department of Computer Science
University of Ottawa
Ottawa K1N 6N5, CANADA

The modern theoretical computer science literature is preoccupied with efficient data structures to code and store ordered sets. Among these, graphical ones play a decisive role. Loosely speaking, graphical data structures must be DRAWN in order that they may be easily READ.

Besides the well known metaphors inspired by layout design, project management, and database design, several unexpected application areas are driving our recent investigations. These include ice flows, fisheye techniques, molds, and quantum topology.

From the viewpoint of graph drawing there is a common thread to these themes. Each involves an ordered set and each views the ordered set monotonically on a two-dimensional surface. We are led ineluctably to upward drawings on an orientable surface in 3-space. Its edges are drawn monotonically with respect to the south-north direction.

The tools appropriate to this analysis come from topological graph theory and differential topology. In turn, this work leads to new algorithmic considerations. How difficult is it to pentangulate a covering graph? Does every ordered set have a cellular upward drawing? How to approximate the critical points of a surface with an upward drawing of the same genus?

Second International Conference on Ordinal Data Analysis. University of Massachusetts, Amherst, MA 01003. October 15-17, 1993

The Meaningfulness of Ordinal Comparisons for General Order Relational Systems

Fred S. Roberts

Department of Mathematics
Rutgers University
New Brunswick, NJ 08903-2101, USA

One of the most useful concepts of measurement theory is that of meaningfulness or invariance. In general terms, a statement using scales of measurement is called meaningful if its truth or falsity is unchanged if all scales in the statement are replaced by other acceptable scales. The theory of meaningfuless has been applied to index numbers, averaging and merging procedures, combinatorial optimization, psychophysical scaling, and analysis of order and matching experiments, to name just a few of its applications. The simplest statement whose meaningfulness needs to be understood is the ordinal comparison $f(a) > f(b)$, the assertion that item a's assigned scale value is larger than item b's. In this talk, we describe the systematic attempt to understand when this ordinal comparison is meaningful. The talk will concentrate on the case when the scale arises from a homomorphism f into a general order relational system, a union of m-ary relations induced on the set of real numbers by rankings of the integers in $\{1, 2, \ldots, m\}$. General order relational systems arise from the concept of m-point homogeneity which plays a central role in the theory of scale type and meaningfulness. The talk will explain the connection, and then will describe how to determine if a homomorphism into a general order relational system leads to meaningful ordinal comparisons.

Second International Conference on Ordinal Data Analysis, University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

TOSCANA'1
A Management System for Conceptual Data

Tammo Rock

Fachbereich Mathematik
Technische Hochschule Darmstadt
Schlossgartenstr. 7
64289 Darmstadt, GERMANY

TOSCANA'1 is a computer management system for conceptual data processing. It allows to implement and treat conceptual data systems built on large data contexts and data bases. Conceptual data systems (as they are developed in the research group on formal concept analysis at the TH Darmstadt) visualize conceptual structures inherent in the data by nested line diagrams. This method of knowledge representation can be effectively used for focussing and browsing in large conceptual structures. TOSCANA'1 manages these procedures by so-called conceptual scales. The user can choose those scales according to his interests. By examples, it will be demonstrated how to use interactively TOSCANA'1 for different purposes.

References:

P. Scheich, M. Skorsky, F. Vogt, C. Wachter, R. Wille: *Conceptual Data Systems*. In: O. Opitz, B. Lausen, R. Klar (eds.): Information and Classification. Springer, Berlin–Heidelberg 1993, 72-84

M. Skorsky, R. Wille: *Zwischenbericht des Projektes Begriffliche Wissensverarbeitung auf dem Gebiet des Baurechts und der Bautechnik*. TH Darmstadt, 1993.

Simulation Based Inference in Econometrics

Til Schuermann

AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974, USA

New techniques in simulation-based inference are applied to joint test of adaptive, rational and naive expectations using categorical business survey data. We assume that the survey responses for expectations and realizations are triggered by jointly normally distributed continuous latent variables which allows for the construction of a simple likelihood function. However, this simple likelihood contains multi-(four)dimensional integrals and requires simulators to evaluate. Simulated maximum-likelihood estimation is carried out using the smooth recursive conditioning (SRC) simulator due to Geweke, Hajivassiliou, Keane and McFadden, a method which is consistent and has low variance. The latter is crucial when maximizing the log-likelihood directly by simulation. Identification of the parameters is achieved by placing restrictions on the response thresholds and/or the variances. We apply this method to micro-level business survey data from Switzerland and the United Kingdom. We find that we can reject *all three* hypotheses.

Implicit Knowledge Structures of Psychoanalysts Concerning the Suitability of Patients for Psychotherapy

Norbert Spangenberg

Sigmund-Freud-Institut
Mylinstr. 20
6000 Frankfurt, GERMANY

Psychoanalytic knowledge is based on highly abstract theoretical assumptions. Nevertheless it is a practical knowledge which helps the analyst to understand his patient, to steer the therapeutic process by his interventions and to cope with the psychic charges of the emotionally dense interaction process. I want to show how the knowledge structure of experienced psychoanalysts was explored. The question is how the (inflexible) abstract knowledge interferes with practical knowledge which must be highly flexible because every patient is different.

Second International Conference on Ordinal Data Analysis. University of Massachusetts. Amherst. MA 01003. October 15-17. 1993

Weight of Evidence for Strict Simple Ordering in Pair Comparison Data Subject to Error

John P. Sutcliffe

Department of Psychology
University of Sydney
Sydney, NSW 2006, AUSTRALIA

There is a wide-spread interest and utility in "putting things into rank order". For that purpose, the method of pair comparisons is to be preferred to the intuitive method of rank order, because the latter imposes orderability regardless, sometimes spuriously, whereas the former allows separation of non-orderable from orderable domains and yields rank orders only in the latter appropriate cases.

As noted, the method of pair comparison allows testing of the hypothesis of the orderability of a domain. In particular, transitivity is necessary for simple ordering. One might choose to reject the hypothesis of orderability on the evidence of intrasitivity provided by observed circular triads. On the other hand, if the given domain were presumptively simply orderable, to save the main hypothesis one could entertain a supplementary hypothesis that the circular triads had arisen from errors of observation. If one could identify such "errors", then one could "correct" them to remove any intransitivities, and thence establish a simple order.

This paper addresses the question of how to deal with circular triads via a theory of errors of observation. If one allows "error" then in general any one of the pair comparisons made may be in error, one can obtain any one of $n!$ orderings of n distinct objects. A "corrected solution" in such a case is arbitrary. More restrictively, however, one can use "weight of evidence" to rule out error for some pair comparisons and to indicate otherwise where the errors may lie. An algorithm, justified by several theorems of the theory, serves to partition the space of data possibilities. There are five main categories of data, one of which is for error free transitivity, another of which is for intractable intransitivities, and the others cover various degrees of amenability to unique "correction" to full transitivity.

Conceptual Data Systems as a Tool for Ordinal Data Analysis

Rudolf Wille

Fachbereich Mathematik
Technische Hochschule Darmstadt
Schlossgartenstr. 7
64289 Darmstadt, GERMANY

Data analysis always uses concepts to elaborate interpretations of given data. Therefore, a formal treatment of (ordinal) data has to include some formal understanding of concepts and conceptual relations. Formal Concept Analysis yields such understanding by mathematizing the philosophical view of a concept as a unit of thought constituted by its extension and its intension (comprehension). The advantage of Formal Concept Analysis is that it unfolds the inherent conceptual structures of the data without data reduction; furthermore, it allows an informative visualization of such conceptual structures by line diagrams.

Ordinal data can be effectively analyzed by methods of Formal Concept Analysis; this will be demonstrated by examples. Even larger data sets may be conceptually represented by using nested line diagrams. Recent experiences with applications in different areas have shown the need of Conceptual Data Systems which provide an even better treatment of large data contexts which, in particular, allow focussing and browsing in conceptual structures. A model for Conceptual Data Systems has been developed during the last three years including a computer management system named TOSCANA'1. First promising applications show that implemented Conceptual Data Systems will be a useful tool for Ordinal Data Analysis.

References:

P. Scheich, M. Skorsky, F. Vogt, C. Wachter, R. Wille: *Conceptual Data Systems.* In: O. Opitz, B. Lausen, R. Klar (eds.): Information and Classification. Springer, Berlin–Heidelberg 1993, 72-84

F. Vogt, C. Wachter, R. Wille: *Data analysis based on a conceptual file.* In: H. H. Bock and P. Ihm (eds.): Classification, data analysis, and knowledge organization. Springer, Berlin–Heidelberg 1991, 131-140

R. Wille: *Concept lattices and conceptual knowledge systems.* Computers & Mathematics with Applications 23 (1992), 493-515.

R. Wille: *Begriffliche Datensysteme als Werkzeug der Wissenskommunikation.* In: H. H. Zimmermann, H.-D. Luckhardt, A. Schulz (eds.): Mensch und Maschine — Informelle Schnittstellen der Kommunikation. Universtätsverlag Konstanz, Konstanz 1992, 63-73.

Geometric Representation of Ordinal Contexts

Uta Wille

FB Mathematik
Universität Giessen
Arndtstr. 2
Giessen, GERMANY

Ordinal contexts are given by data matrices representing attribute values for certain objects where the values of each attribute form an ordered set. A *geometric representation* of an ordinal context is understood as the combination of an object embedding into an ordered vector space, an attribute embedding into its dual space, and an attribute value embedding into the ordered field of scalars such that an object g has the value w of the attribute m iff $\varphi_m(x_g) = r_w$ for the corresponding vector x_g, the linear form φ_m, and the scalar r_w (cf. [2]). In 1964, D. S. Scott gave a first characterization of such geometric representation in an n-dimensional vector space over the reals for finitely many objects and $n+1$ attributes (cf. [1]). Since Scott's axiomatization is not finite, it cannot be used to decide whether a geometric representation exists for a given data context. Nevertheless, following Scott's approach, a *representation theorem* can be proved which guarantees the representability of a finite ordinal context by a finite number of axioms if the values of each attribute are distinct. An example will clarify the content of the representation theorem and further questions concerning geometric representation of ordinal contexts.

References:

[1] P. Suppes, D. H. Krantz, R. D. Luce, A. Tversky: *Foundation of measurement.* Vol. 2. Academic Press, San Diego 1989.

[2] U. Wille: Eine Axiomatisierung bilinearer Kontexte. Mitt. Math. Sem. Giessen **200**, 1991, 72-112.

Second International Conference on Ordinal Data Analysis, University of Massachusetts, Amherst, MA 01003, October 15-17, 1993

Growing a Classification Tree Using the Apparent Misclassification Rate

Elias Zintzaras

Laboratory of Mathematical Biology
National Institute for Medical Research
London, ENGLAND

A method to determine the size of a classification tree is proposed. This method is based on the change of this apparent misclassification rate (AMR) of the tree at each growing stage. The method is simple and fast compared to the other classification tree methods which are based on minimizing a cost complexity function (Breiman *et al.*, 1984). To test the method, it was used to classify species of fungi, and the results are in good agreement with those obtained by linear discriminant analysis. Also, 21 proteins with known structures and functions were classified using the proposed method. For this purpose the coefficient of variation (CV) for several properties of the secondary structures of these proteins has been used. Again, the results were in good agreement with the classification obtained by Orengo *et al.* (1992) using dynamic programming.